

Tracking People Across Disjoint Camera Views

A Thesis Submitted For the Degree of Doctorate of Philosophy


By

Christopher Madden

Faculty of Information Technology
University of Technology, Sydney
Australia

July 3, 2009

I, **Christopher Madden**, certify that the work in this thesis titled “**Tracking People Across Disjoint Camera Views**” has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text. I also certify that the thesis has been written by myself. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. The undersigned certify that they have read this thesis and that in their opinions it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

A handwritten signature in black ink, appearing to read 'Chris Madden', with a long horizontal flourish extending to the right.

Christopher Madden

Date: July 3, 2009

Principal Supervisor: Massimo Piccardi

Acknowledgements

I would like to thank all those people who have been involved in some way with the creation of this thesis. My biggest thanks goes to Professor Massimo Piccardi for being my supervisor and assisting me with learning all of the areas I have needed to know for this work. And for filling me in on the areas that I still haven't mastered yet. To my loving wife and family thank you for all the caring support. I would like to promise that I will talk a little less about video surveillance from now on; however we both know that is unlikely to happen. I would also like to thank my friends and colleagues who have assisted me throughout my candidature. Especially those who have read, or even reread, this thesis and helped me to improve it to its current state. I would also like to thank you, the reader, for your time as this thesis is not only written to explore my knowledge of the area, but hopefully to pass some of that on to you.

Abstract

Tracking people around surveillance systems is becoming increasingly important in the current security conscious environment. This thesis presents a framework to automatically track the movements of individual people in large video camera networks, even where there are gaps between camera views. It is designed to assist security operators, or police investigations by providing additional information about the location of individuals throughout the surveillance area. Footage from an existing surveillance system has been used to test the framework under real conditions. The framework uses the similarity of robust shape and appearance features to match tracks. These features are extracted to build an object feature model as people move within a single camera view, which can be compared across cameras. The integration of matching similarities in the temporal domain increases the robustness to errors of many kinds. Frames with significant segmentation errors can be automatically detected and removed based upon their lack of similarity to the other models within the same track, increasing robustness.

The shape and appearance features used to generate the object models are based upon features humans habitually use for identifying individuals. They include a height estimate, a Major Colour Representation (MCR) of the individuals global colours, and estimates of the colours of the upper and lower portions of clothing. The fusion of these features is shown to be complementary, providing increased discrimination between individuals. The MCR colour features are improved through the mitigation of illumination changes using controlled equalisation, which improves the accuracy in matching colour under normal surveillance conditions and requires no training or scene knowledge. The incorporation of other features into this framework is also relatively straightforward.

This track matching framework was tested upon four individuals across two video cameras of an existing surveillance system. Existing infrastructure and actors were used to ensure that ground truth is available. Specific cases were constructed to test the limitations of the system when similar clothing is worn. In the data, the height difference ranges from 5 to 30 centimetres, and individuals may only be wearing 50% of similar clothing colours. The accuracy of matching an individual was as high as 91% with only 5% false alarms when all the system components were used. This may not become a fully automated system, but could be used in semi-automated or human assisted systems, or as the basis for further research into improved automated surveillance. Application areas range from forensic surveillance to the matching of the movements of key individuals throughout a surveillance network and possibly even target location.

Contents

1	Introduction	1
1.1	Aim	4
1.2	Scope	5
1.3	Contribution	6
1.3.1	Publications	7
1.4	Thesis Overview	8
2	Literature Review	10
2.1	Motion Detection and Object Segmentation Techniques	10
2.1.1	Background Modelling	12
2.1.2	Removal of Shadow and other Segmentation Noise	16
2.2	Colour Space Research	18
2.3	Statistical Similarity Measurements	25
2.4	Object Tracking	27
2.4.1	Complexity Factors within Object Tracking	30
2.4.2	Feature-Based Tracking	31
2.4.3	Model-Based Tracking	32
2.4.4	Mean Shift-based Tracking	35
2.4.5	Summary of Tracking Literature	36
2.5	Object Classification	37
2.6	Current Disjoint Camera Tracking Methods	38
2.7	Literature Summary	47
3	Colour-based Robust Appearance Features	49
3.1	Appearance Feature Background	50
3.2	MCR Colour Feature Extraction	53
3.2.1	Optimising MCR Using an Online k-means Algorithm	56
3.3	Improving Robustness Using Incremental MCRs	58
3.4	Comparing MCR or IMCR Appearance Features	60
3.4.1	Time Integration of Similarity	63
3.5	Extracting Spatial MCR Colour Features	65
3.6	Experimental Validation of MCR Appearance Features	68
3.6.1	Colour Experiments on Manually Segmented Individuals	69
3.6.2	Colour Experiments on Automatically Obtained Tracks	72
3.7	Discussion of MCR Appearance Results	77
3.8	Summary of MCR Appearance Features and Future Enhancements	79

4	Mitigating the Effects of Changes in Illumination	82
4.1	Illumination Mitigation Background	83
4.2	Illumination Filtration	87
4.3	Histogram Stretching	88
4.4	Histogram Equalisation	90
4.5	Comparing Illumination Mitigation Techniques	93
4.6	Experimental Comparison of Mitigation Techniques	96
4.7	Discussion of Illumination Mitigation	100
4.8	Summary of Illumination Mitigation and Future Enhancements . .	101
5	Identification of Segmentation Errors	104
5.1	Segmentation Error Identification Background	105
5.2	Identifying Segmentation Errors Through Changes in Bounding Box Height	106
5.3	Identifying Segmentation Errors Through Appearance Feature Anal- ysis	107
5.3.1	Comparing Colour Features Between Frames	111
5.3.2	Typical MCR Patterns of Major Segmentation Errors . . .	112
5.4	Experimental Validation for the Identification of Major Segmen- tation Errors	114
5.5	Discussion of Segmentation Error Identification	116
5.6	Summary of Segmentation Error Identification and Future En- hancements	117
6	Height Based Robust Shape Feature	119
6.1	Shape Feature Background	120
6.2	Obtaining Height Estimates Using Camera Calibration	124
6.3	Improved Automatic Monocular Height Estimates	128
6.4	Statistically Comparing Height Features	132
6.5	Experimental Verification of Height Estimation	134
6.5.1	Height Experiments Comparing Manual and Automatic Height Estimates	134
6.5.2	Height Experiments Using a Larger Dataset	137
6.6	Discussion of Height Results	139
6.7	Summary of Height Feature and Future Enhancements	140

7	Fusion Methods and Results for Combining Robust Features	142
7.1	Classifier-based Fusion Background	143
7.2	Classifier-based Fusion for Integrating Features across Differing Time Scales	146
7.3	Results From Fused Features	147
7.3.1	Evaluation of the Statistical Models	148
7.3.2	Evaluation of Fusing Features	150
7.4	Summary of Fused Features	153
8	Conclusions	156

List of Figures

1	Approaches incorporating spatial colour information	52
2	Major Colour Representation of ‘tn_flower’	56
3	Original ore gold rose image (left) and reprojection of the 90% most frequent pixel clusters after 7 k-means iterations (right) . . .	57
4	MCR changes for 20 most significant colours with iterations of the k-means optimisation	57
5	MCR from three automatically detected people	59
6	IMCR matching of two tracks using time integration	64
7	Examples of upper and lower regions of segmented individuals . .	67
8	Same individuals observed in camera 3a and camera 5	69
9	Differing individuals observed in camera 3a and camera 5	71
10	Typical frames used for test cases	72
11	Typical backgrounds used for test cases	73
12	Four people of interest (Person’s A, B, C, D from left) and good automatically segmented masks (from frames 775, 1095, 1542, 2044)	74
13	Poor segmentation in two sample cluttered frames	76
14	Accuracy of individual colour features	77
15	Sample people of interest and their red histograms under differing illumination conditions	83
16	Individuals R values before and after illumination filtration	88
17	Histogram stretching of the individual’s pixels	90
18	Full equalisation of the individual’s pixels	91

19	Controlled equalisation of the individual's pixels with varying k values	92
20	Intersection of the $H0$ and $H1$ curves of the height feature	95
21	Example of changes in bounding box height ratios where a large segmentation errors occurs and 5 sample frames from the track including the erroneous frame	108
22	Example of upper and lower regions from three segmentations of one person	110
23	Example of upper and lower regions from three segmentations of a second person	110
24	Three typical error patterns of frame based pairwise similarity comparisons given between 0 and 1	113
25	Four people of interest and automatically segmented masks of good quality	114
26	ROC curves of the height, colour and fused feature results	115
27	Accurate location of the bottom point improves height estimates	125
28	Manually identified key image points for the track of one individual	127
29	Curvature values for a single individual showing curvature key points beginning at the top left most object pixel	130
30	Height estimates for 5 tracks demonstrating the manual height estimates	132
31	An individual's key points and height estimates using automatic and manual techniques on the poorly segmented track 13	135
32	ROC curves of the height feature results	138
33	Example $P(s_{UC} H0)$ and $P(s_{UC} H1)$	145
34	Four people of interest	148
35	ROC curves of the height, MCR and fused feature results	151
36	ROC curves for fusing spatial colour MCR's, fusing all the colour MCR's, and fusing all the features	152
37	Pictorial storyboard summary of potentially matching tracks	155

List of Tables

1	Results of IMCR Matching - same person	70
2	Results of IMCR Matching - differing people	72
3	Results of IMCR matching - differing people	73
4	Results of automated IMCR matching - 6 different cases	75

5	Global similarity measurements for matching and non-matching tracks	97
6	Upper MCR similarity for matching and non-matching tracks . . .	98
7	Lower MCR similarity for matching and non-matching tracks . . .	100
8	PD and PFA values of Bounding Box and MCR features for detecting segmentation errors	115
9	Auto height estimates of a 1710 mm individual over 15 tracks . .	136
10	Ground Truth of Participants	148
11	How variations to the optimum threshold affect% error rates . . .	149